How Valid Is The DIBELS Oral Reading Fluency Test as an Indicator for Identifying At-Risk Readers and Documenting Their Progress?

A Research Project Report

Presented to

The Graduate Faculty

University of Wisconsin-Superior

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science in Education-Reading

December 2012

By

Tim Foslin

How Valid is the DIBELS Oral Reading Fluency Test as an Indicator for Identifying At-Risk Readers and Documenting Their Progress?

Tim Foslin

A Research Project Report

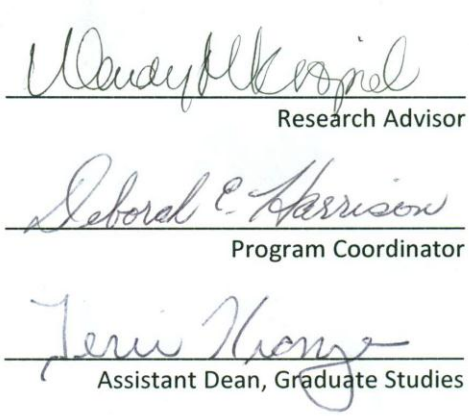


Table of Contents

Chapter 1

Introduction

The passage of the federal No Child Left Behind (NCLB) legislation in 2001, with its emphases on accountability, quality of instruction, and data-driven decision making, has created a national debate centered around the efficacy of a variety of practices in our nation's schools. The effective use of valid and reliable quantitative assessments to inform instructional and curricular decisions made by reading practitioners and decision makers at all levels is vitally important if student achievement is to be maximized. The focus of this research will be on two of these assessments which were used at the Lac Courte Oreilles (LCO) Ojibwe School from 2004 – 2010 as part of the implementation of the federally funded Reading First program - the Dynamic Indicators of Basic Early Literacy (DIBELS) and the Stanford Achievement Test (SAT).

Problem Statement

In response to the mandates of the NCLB legislation, the International Reading Association (IRA) outlined six principles to guide practitioners in carrying out what is generally referred to as Response to Intervention (RTI): effective instruction provided by classroom teachers, differentiated instruction tailored to meet individual student needs, collaboration between staff members to provide high quality interventions for struggling readers, use of a systemic and comprehensive approach K-12, high quality professional development to increase expertise of all staff members, and lastly, the one which directly relates to this research – the use of assessment data from a variety of sources to inform instructional and intervention decisions (IRA, 2008).

However, in their implementation of these principles, many reading practitioners have tended to disregard the NCLB legislation's prescription relating to the use of instructional methods based on Scientifically Based Reading Research (No Child Left Behind Act, 2002). The whole language approach (more recently referred to as balanced literacy) is favored instead, giving classroom teachers a wider latitude and the prerogative to make instructional decisions as they see fit (Moats, 2000). These decisions are based on professional judgment, personal preference, or any number of other criteria, as opposed to following the dictates of what SBRR prescribes (Moats, 2000). There is ongoing debate concerning these issues, not the least of which is how to effectively choose and make use of valid and reliable quantitative assessments to help inform and drive instructional and curricular decision making.

Many educators who make use of the balanced literacy approach in their classrooms are often reluctant to fully support the use of two types of quantitative measures: 1) standardized tests - summative measures used to evaluate instructional effectiveness, and 2) screening tests - formative assessments used to help teachers identify struggling readers and make the necessary adjustments to instruction (Dessoff, 2007). Yearly standardized testing of students has been a practice employed by many districts in our country, sometimes voluntarily, and at other times in response to state or federal accountability mandates. In the view of educators who support balanced literacy, reinforced by the International Reading Association's aforementioned guiding principles on Response to Intervention, the data from these assessments lacks validity due to the use of contrived texts, tasks generated specifically for assessment purposes, and activities that don't reflect authentic literacy (IRA, 2008). Additionally, in the last few decades, there has been an marked increase in the number of school districts using another type of quantitative assessment, screening tests, to help in the early identification of struggling readers. The

Dynamic Indicators of Basic Early Literacy (DIBELS) is one of the most common measures employed for this purpose (Good & Kaminski, 2002). This rise in the use of screening measures has also been accompanied by growing concern among the balanced literacy educators, regarding both their administration and the use of data generated by them (Pearson, 2007; Goodman, 2007).

In part to satisfy the requirements mandated by the NCLB federal legislation and the Reading First initiative's guidelines, the LCO School used both the DIBELS screening and progress monitoring assessments as well as the Reading First Stanford Achievement Test (SAT) on a yearly basis, from 2004 – 2010. The DIBELS assessments, which directly measure the three foundational components of the reading process (phonemic awareness, phonics, and fluency), are designed to be predictive of future reading success, not only in these three areas, but in relation to the overall reading process, including vocabulary and comprehension as well (Good & Kaminski, 2000). One goal of this research was to find out if the results of the DIBELS assessments, administered three times yearly to the K-3rd grade students the LCO School, which measure specific skills, correlated to reading achievement in general (and thus, are predictive of overall reading success) as measured by the yearly administration of the Reading First SAT. If so, this will be one more piece of evidence to help substantiate the validity and use of the DIBELS Oral Reading Fluency test data to identify and document the progress of struggling readers.

Rationale for Study

The achievement gap between white students and their ethnic minority peers in our nation's schools has been well documented (Willingham, 2012). All of the students served at the LCO School from 2004 - 2010 were of Native American descent. In addition, all students fell under the federal government's definition of low socio-economic status – participation in the free and reduced lunch program (Euneau, 2004 – 2010). As minority children growing up in poverty, these students, and the staff at the LCO School whose mission it was to educate them, had many obstacles to overcome and overwhelming odds to face in their quest to succeed in reading, school in general, and life after graduation.

From 2004 – 2010, when the LCO School received funding for the federal Reading First grant, the staff made great strides towards accomplishing the program's goal of having all students reading at or above grade level by the end of third grade. The school met or exceeded the criteria established for Annual Yearly Progress (AYP), as mandated by the federal NCLB legislation in each of those years (Euneau, 2004 – 2010). In addition, during the six years of program implementation, there were no placements of K-3rd grade students into special education as a result of a reading disability (Mueller, 2004 – 2010). Staff members were clearly starting to make the paradigm shift from the more passive, developmental readiness, "wait to fail" approach, advocated by proponents of balanced literacy, which resulted in over-identification of students into special education, to the more proactive, early identification, response to intervention, SBRR approach of the Reading First program. This research is aimed at supporting an increased fidelity to these principles and practices – specifically fostering among staff members a more widespread understanding, acceptance, and use of DIBELS

screening assessments and data to identify and adjust instruction more effectively for at-risk readers.

Hypothesis

There will be a high level of correlation between the 1st – 3rd grade student scores on the DIBELS Oral Reading Fluency (ORF) spring benchmark assessments and their Total Reading Scores on the Reading First SAT from 2004 – 2010. This evidence will help substantiate the validity and use of the DIBELS test data to identify and help struggling readers. If the DIBELS ORF test can reliably predict performance on the Reading First SAT, teachers can then use the DIBELS screening data as a basis or starting point to look for early warning signs of reading struggles, and design research-based instruction and interventions to help the at-risk readers succeed (Baker et al. 2008).

Summary of Study

My subjects were 190 students enrolled in the 1st – 3rd grades at the LCO School from 2004 – 2010. All of the children were Native American who lived within the boundaries of the federally recognized Lac Courte Oreilles Ojibwe reservation. Both boys and girls were included in the study.

All of the students participated in the free and reduced federal school lunch program. By virtue of their ethnic minority and low socio-economic status, all of the students in the study were at risk for developing or already having deficits in one or more of the foundational areas (phonemic awareness, phonics, fluency, vocabulary, comprehension) involved in the reading process (Ladson-Billings, 2006). Described by their scores on the DIBELS benchmark

assessments, students needing all levels of assistance to succeed in reading were included in the study (intensive – needing a substantial amount of assistance; strategic – needing extra assistance; benchmark – needing little/no extra assistance). The study included both readers who were already identified with reading (or other) disabilities, having Individual Education Plans (IEPs) developed and in place, as well as those who, though struggling, had not been identified as having a disability.

I analyzed the data from two assessment sources given each year to the 1st – 3rd graders between 2004 - 2010:

1. DIBELS Oral Reading Fluency (ORF) spring benchmark testing: This testing was administered three times per year – fall (early September), winter (early December – mid January), and spring (early May) - to all K-3rd grade students. Trained assessors were contracted through the consulting firm, EndVision, which specialized in data management and statistical analysis for schools. Teams of assessors came in to administer the fall and spring DIBELS testing, in an effort to increase the validity and reliability of the results. Some students took the tests with approved accommodations, as specified in their IEPs. A reading coach from a nearby Bureau of Indian Education (BIE) Reading First school administered the winter benchmark assessments to the students.
2. Reading First Stanford Achievement Test (SAT) - 10th Edition: This testing was sdministered every year in early May to the K-3rd grade students. It was administered to classroom groups by their teacher, spread out over two to four morning testing periods during the week. As with the DIBELS tests, some students were administered these assessments with approved accommodations, as specified in their IEPs. The results were

mailed in to the publisher, NCS Pearson, Inc., to be machine scored, the results being sent back typically within a two month time frame.

I disaggregated the data according to the following subgroups within the 1st - 3rd grade population:

1. Grade Level Groups
2. Classroom Groups
3. Gender Groups
4. Groups Based on DIBELS Cut Scores
5. Groups Based on Reading First SAT Cut Scores

Limitations and Delimitations of Study

This research encompassed a sampling of 190 1st – 3rd grade students as subjects. Even though it was a relatively small sample, they were, on the other hand, fairly homogeneous – all were Native Americans from low socio-economic status homes participating in the Reading First program. There was also a certain amount of mobility among the students and families who attended the LCO School, with transitions occurring during each year as well as across school years. Some left and came back within the same year, while others were first time arrivals to the school, while still others left and didn't return. For the purposes of this study, only students in a given school year that started and ended the year at the LCO School (and participated in both the DIBELS benchmark and the Reading First SAT administrations) were included in the analysis. The Hawthorne effect should be low, due to students being exposed to both testing situations, the DIBELS ORF Subtests and the Reading First SAT, every year starting with 1st grade.

Referenced Definitions of Terms

1. Phonemic Awareness – The ability to notice, think about, and work with the individual sounds in spoken words. Before children learn to read print, they need to become aware of how sounds in words work. They must understand that words are made up of speech sounds, or phonemes (National Institute for Literacy, 2000).
2. Phonics Instruction – Teaches children the relationships between the letters of written language (graphemes) and the individual sounds of spoken language (phonemes). It teaches children to use these relationships to read and write words (National Institute for Literacy, 2000).
3. Reading Fluency – The ability to read a text accurately and with prosody. When fluent readers read silently, they recognize words automatically. They group words quickly to help them gain meaning from what they read. Fluent readers read aloud effortlessly and with expression. Their reading sounds natural, as if they are speaking (National Institute for Literacy, 2000).
4. Vocabulary – The words we must know to communicate effectively. Oral vocabulary refers to words we use when speaking or recognize when listening. Reading vocabulary refers to words we recognize or use in print (National Institute for Literacy, 2000).
5. Reading Comprehension – The reason for reading. If readers can read the words but do not understand what they are reading, they are not really reading. Good comprehenders have a purpose for reading and think actively as they read (National Institute for Literacy, 2000).

6. DIBELS Recommended Levels of Support (Cut Scores)
    a. Intensive – Students at high risk of not reaching the DIBELS benchmark goal during the next testing period (Good, Simmons, Kame'enui, Kaminski, Wallin, 2002).
    b. Strategic – Students at some risk of not reaching the DIBELS benchmark goal during the next testing period (Good, Simmons, Kame'enui, Kaminski, Wallin, 2002).
    c. Benchmark – Students at low risk of not reaching the DIBELS benchmark goal during the next testing period (Good, Simmons, Kame'enui, Kaminski, Wallin, 2002).
7. Fidelity of Reading Program Implementation – consists of two components:

    1) Content – accurately teaching the reading program the way it was designed; and

    2) Delivery – using effective teaching techniques and strategies to maximize student learning (Brenner & Hiebert, 2010).
8. Native American Student – A student who has a large enough "blood quantum quota" (documented Native American ancestry) to be considered part of a tribe and is permitted to enroll at the LCO School.
9. Annual Yearly Progress (AYP) – Under the provisions of the federal No Child Left Behind (NCLB) legislation, an individual district's measure of progress toward the 2014 goal of 100% of students performing up to grade level academic standards in the core subjects of Reading/Language Arts, Math, Science, and Social Studies. (No Child Left Behind Act, 2002).
10. Test Validity – The degree to which a test measures what it purports to measure (McKenna, M., & Stahl, S., 2003).
11. Test Reliability – The consistency of a measurement, or the degree to which an instrument measures the same way each time it is used, under the same conditions, with

the same subjects. In short, it is the repeatability of a measurement or its general dependability. A measure is considered reliable if a person's score on the same test given twice is similar (McKenna, M. & Stahl, S., 2003).

Conclusion

This study is an attempt to shed some light on the assessment practices used at the LCO School in the context of the No Child Left Behind (NCLB) legislation, the Reading First program, Response to Intervention (RTI) recommendations, and Scientifically Based Reading Research (SBRR) guidelines. Ascertaining the extent which the DIBELS ORF spring screening test results, used as a specific measure of fluency skill to identify at-risk readers, serve as an accurate predictor of a student's overall score on the Reading First SAT, as an indication of general reading achievement and used as an outcome and accountability instrument, will hopefully result in more staff members accepting the DIBELS ORF as a valid indicator to identify at-risk readers and document their progress, accompanied by a more widespread and effective use of the DIBELS assessment data.

Chapter 2

Introduction

Formative assessments, both for screening and progress monitoring purposes, play an important role in identifying struggling readers for intervention and in guiding the decisions of both the classroom teacher and specialist to carry out effective instruction. The validity and predictability of one such battery of measures, The Dynamic Indicators of Basic Early Literacy (DIBELS) is important to establish for school staff members to fully support and implement data-driven instructional decision making. Subjective teacher judgments of objective assessment data are a factor influencing the quality of this decision-making process. In addition, even though a district can have many of these components in place, their effectiveness will be limited unless they are co-ordinated in a comprehensive, integrated system of support.

The Role of Assessment in the Identification of At-Risk Readers and Design of Appropriate Interventions

Even though there is progress being made in our nation's schools towards the Reading 1st goal of having every child reading at grade level by the end of third grade, there is still much work to be done. Currently, even with research-based methods being applied, both by classroom teachers and specialists, there still remains a small percentage, 2 – 12%, of the total school population, that continues to need intensive levels of intervention. They have been referred to by one researcher as "treatment resisters" or "non-responders to treatment" (Torgesen, 2000). Due in part to the increasing tendency in schools towards the over-identification of children for

special education services, the NCLB federal legislation has mandated that school professionals be required to demonstrate unresponsiveness to both primary and secondary interventions instead of the commonly used discrepancy between scores on IQ and achievement tests when qualifying students for special education services (Torgesen, 2000).

Although underlying issues such as attention/behavior problems, language limitations, orthographic processing weaknesses, naming-speed deficits, parental level of education, and socio-economic status can be present, the two most pervasive and universal characteristics that struggling readers display (both at-risk and learning disabled) are phonological processing deficits and inadequate word reading skills (Al Otaiba & Fuchs, 2002). The presence of either of these deficits can cause a child to develop reading dysfluency, negatively affecting comprehension.

It has also been found that fewer students were unresponsive to instruction and intervention that included the direct teaching of phonics and decoding skills (Al Otaiba & Fuchs, 2002). Other studies found individual tutoring to be most promising for reducing unresponsiveness to intervention, even in the presence of an evidence-based, class-wide reading program and multi-phased, phonological/phonics focused interventions, which, although being good first steps, are not sufficient for helping students who continue to non-respond be successful (McMaster et al, 2005; Vadasy et al, 2002).

The Role of Assessment in a Multi-Tiered System of Support

Researchers point out, that through their intervention efforts, school personnel should not expect to ameliorate the symptoms presented by struggling readers in a relatively short period of time. Indeed, these targeted strategies, skillfully implemented, need to be applied, in some cases,

over a period of years, instead of months. In spite of this necessity for ongoing, intensive, long-term intervention, it has been clearly demonstrated that early prevention efforts, in kindergarten, first grade, or earlier, reduce the time required for remediation (Vaughn et al, 2003).

School districts that are adopting system-wide, prevention-focused approaches to reading instruction and intervention have a higher probability of helping struggling readers succeed (Chard et al., 2008; Dwyer, Osher, & Hoffman, 2000; Haager et al., 2007; National Research Council, 1998; Simmons et al., 2003). The key components in this type of effective system include (a) prioritized instruction; (b) evidence-based instructional practices; (c) coordinated, integrated, and differentiated multi-tiered interventions; (d) utilization of screening and progress monitoring assessments; (e) data-driven decision making focused both on individual student and group levels; and (f) sustained, targeted professional development (Haager et al., 2007; Vaughn, Linan-Thompson, & Hickman, 2003).

Instruction that is not well designed or of sufficient intensity can cause reading difficulties and even intensify existing problems (Chard & Kame'enui, 2000). Vellutino et al. (1996) found that many students identified as having a reading disability and placed in special education programs were, in reality, struggling readers without disabilities who had not experienced effective instruction from their prior year's teachers. This lack of effective instruction, an environmental factor, was as much a cause for their reading struggles as any innate characteristics or biological factors (Vellutino, Scanlon, Small, & Fanuele, 2003).

Data-Driven Instructional Decision Making

Solid, research-based reading instruction and intervention requires that many pieces be put into place within the culture of an elementary school for effective results to take place (National

Institute of Child Health and Human Development, 2000). The primary components of direct, explicit, and systematic, classroom-based reading instruction need to be in place. This daily instruction needs to be implemented in skillful ways by a teacher knowledgeable in all facets of the reading process. Timely, focused, and phonological/phonics based individual and/or small group interventions need to be in place as well (Center for the Improvement of Early Reading Achievement, 2001).

But that list, although seemingly comprehensive, is not all that needs to take place. The implementation of valid and reliable formative assessments, which serve to inform and shape classroom instruction, is one of the most important pieces to consider (Dessoff, 2007). To that end, the Dynamic Indicators of Basic Early Literacy (DIBELS) screening measures (Good & Kaminski, 2002), which focus on the critical early literacy skills necessary for early readers, is one such assessment program many states and districts are currently using for these purposes (Baker et al., 2008). They can be helpful to educators in identifying at-risk readers who are in need of more intensive interventions, as well as tracking how students respond to intervention instruction (Olson, 2007). These early literacy assessments also serve to inform and drive the instruction for classroom teachers of reading, enabling them to build their teaching around and respond to the individual scores and trends they see (Dessoff, 2007).

DIBELS Assessments

The Dynamic Indicators of Basic Early Literacy (DIBELS) assessments are short, one minute, individually administered, standardized fluency measures of the basic skills involved in early literacy, designed to be used with kindergarten – 6th grade students (Good & Kaminski, 2002). The seven subtests are: Initial Sound Fluency (ISF) and Phonemic Segmentation Fluency

(PSF) – measures of phonemic awareness; Letter Naming Fluency (LNF) – a measure of alphabetic knowledge; Nonsense Word Fluency (NWF) – a phonics measure; Oral Reading Fluency (ORF) – a measure of reading speed and accuracy; Word Use Fluency (WUF) – a measure of vocabulary; and Retell Fluency (RTF) – a comprehension measure. These tests can be used for screening purposes, administered three times yearly (fall, winter, and spring) and for progress monitoring, given every 1 – 4 weeks.

These DIBELS screening assessments were devised to help educators identify struggling, at-risk readers, so that appropriate types and levels of supports can then be implemented. They were designed to support efforts at early intervention, in the process helping to prevent reading difficulties from developing into larger deficits as time passes, thereby reducing the need for more intensive remediation efforts. The data derived from these assessments was also intended to help practitioners evaluate the effectiveness of interventions being carried out so that necessary adjustments can be made as part of the Response to Intervention (RTI) model. In addition, the DIBELS tests were created to be used on a systems level, giving administrators information about the effectiveness of supports in place for struggling readers and what changes in resources need to be made, ideally within a multi-tiered system of instructional support.

The DIBELS assessments were based on the earlier research of Dr. Stan Deno and his colleagues at the University of Minnesota in the 1970's – 80's. They developed and tested economical and efficient measures of student progress called Curriculum-Based Measurement (CBM). The development of the DIBELS measures, initially called, *Curriculum Based Measures of Pre-reading Skills (CBM-P),* was led by two researchers at the University of Oregon, Dr. Roland H. Good III, PhD and Dr. Ruth A. Kaminski, PhD, starting in the 1980's. From 1996 to 2001, the Early Childhood Research Institute on Measuring Growth and

Development (ECRI-MGD), funded by the U.S. Department of Education, supported the work being done. From its early stages, continuing on to the present, many researchers have contributed to the work being carried out by the Dynamic Measurement Group (DMG), which markets the DIBELS products, donating its royalty income to the University of Oregon for further research.

Reading First SAT Assessments

The Reading First Stanford Achievement Test is a version of the Stanford Achievement Test – 10th Edition (Harcourt Assessment, 2004). The SAT 10 was developed and first written in 1926 by the Harcourt Educational Measurement publishing company as a norm-referenced standardized measure of student achievement. It was designed as an outcome measure of reading progress comparing an individual student's score with norms developed from representative samples of achievement taken from a nation-wide grade level group. It has been used on a continuous basis since its original development, but more recently, due to the mandates of the federal NCLB legislation calling for increased accountability in our nation's schools, its use has become much more widespread.

The Reading First SAT is a group-administered series of subtests assessing the five essential components of reading identified by the Reading First federal grant legislation – phonemic awareness, phonics, fluency, vocabulary development, and comprehension strategies. The multiple choice formatted subtests are untimed and can be administered to students in either paper/pencil or electronic forms. The results for each subtest, as well as a student's overall test score, are reported in criterion-referenced form, based on one of three suggested levels of support – Needs Substantial Intervention (performance falling below the 20th percentile), Needs

Additional Intervention (performance between the 20th and 40th percentiles), and At Grade Level (performance at or above the 40th percentile). Student scores are also reported using other statistical indicators – national percentile rankings, stanines, scaled scores, and national curve equivalents.

Predictability of DIBELS ORF

One of the most desirable features of the DIBELS tests in general, and the ORF in particular, is its correlation with other standardized measures of reading achievement. This points to its ability to accurately predict future reading outcome indicators.

For a group of 3rd grade students in Tempe, Arizona, the correlation between scores on the spring DIBELS ORF and Arizona Instrument to Measure Standards (AIMS) reading test was positive and moderately large (r = .741) (Wilson, 2005). In terms of proficiency levels, the ORF identified with good accuracy those in the Low Risk category on the AIMS, and was even better at identifying those in the At Risk category.

For a group of Colorado third graders, the ability of the spring ORF score to predict their proficiency level on the Colorado State Assessment Program (CSAP) reading test was also high (Shaw & Shaw, 2002). Based on a DIBELS score of 110 correct words per minute (cut off for a "low risk" recommendation), the ORF was able to correctly predict a proficient/advanced or unsatisfactory/partially proficient score on the CSAP for 74% of the students. Using a DIBELS score of 90 correct words per minute (within the "at risk" recommendation band) resulted in an even higher percentage – correctly predicting 86% of the students.

Researchers found, that for a group of third graders in Florida, there was a significant correlation between their spring ORF scores and Florida Comprehensive Assessment Test –

Sunshine State Standards (FCAT-SSS) reading score (r = .70, p<.001) (Buck & Torgesen, 2002). In terms of proficiency levels, 91% of the students who reached or exceeded the ORF "low risk" cut off score of 110 correct words per minute reached the "adequate performance" level on the reading portion of the FCAT-SSS.

For a group of third graders in North Carolina, researchers found a significant correlation between the DIBELS spring ORF scores and the North Carolina End of Grade Reading Assessment (r = .73) (Barger, 2003). 92% of the students who scored 110 correct words per minute or better (cut off for a "low risk" recommendation) on the DIBELS ORF achieved the "superior mastery" proficiency level on the NC End of Grade Assessment. An even better prediction rate was found for students who scored 100 correct words per minute on the DIBELS ORF (within the "at risk" recommendation band). 100% of those students attained the "consistent mastery" score on the NC End of Grade Assessment.

Predictability and Effects of Teacher Judgments

Besides having valid and reliable screening tests in place, another key factor affecting identification of at-risk readers and the design of effective interventions is the role teacher judgments of student reading ability play in the process.

Hecht and Greenfield compared First Grade teacher ratings of student reading ability to the results of reading-related tests to find out which had more predictive validity on Third Grade levels of reading skills (Hecht & Greenfield, 2001). The results demonstrated that both teacher ratings and reading-related test data were both similarly predictive and highly redundant in regard to student reading achievement in Third Grade. Both methods of identifying struggling readers, although similarly predictive, provide unique information about a student's strengths

and weaknesses, and thus, in the authors' estimation, both should be relied upon in the context of a comprehensive battery of assessments.

Atkinson examined how a group of third- through fifth-grade teachers read and responded to their students' Stanford Achievement Test 10 (SAT 10) data and reported that the process of making objective, data-driven decisions was infused throughout with the teachers' subjectivity (Atkinson, 2012). Many teachers tended to blame students and their families for low test scores, believing it to be the result of being raised in low socio-economic status (SES), minority homes where education is not valued (Hale, 2001; Lynn et al., 2010; Noguera, 2007). This belief system engendered and perpetuated a "culture of low expectations" (Landsman, 2004). Some teachers blamed the test publishers and felt that their minority, low SES students were not represented in the national norms established. They also questioned the authority of policy makers and educational consultants, feeling they were simply coming up with more rules and always checking up on them. Others objected to the way that current accountability policies and standardized testing negatively affects their autonomy to make instructional decisions and maintains a system of "haves" and "have nots" among the students. A final, and small group of teachers reported that their students' low test scores caused them to reflect on how to change their own instructional practices. They didn't react defensively by blaming the students, but proactively looked at how they could change the instruction, intervention, and assessment practices they were using to help students succeed (Atkinson, 2012).

Conclusion

The use of Scientifically-Based Reading Researched (SBRR) methods by regular classroom teachers to effectively present instruction to their students is but a good first step in the right

direction towards helping all students become successful readers. Well designed, targeted, and timely interventions are other pieces of the puzzle for districts to put into place. The frequent, ongoing administration of valid and reliable assessments that probe critical skills young readers need to master, such as the Dynamic Indicators of Basic Early Literacy (DIBELS) benchmark and progress monitoring tests, is yet another component of an effective school-wide reading program. But, even with these elements in place, success is not guaranteed. A systemic approach, dynamically integrating and orchestrating the individual pieces together into a comprehensive, K-12 program is perhaps needed most of all.

Chapter 3

Introduction

The use of valid and reliable reading assessments to inform instruction, identify children in need of intervention, and gauge the results of the research-based instruction being carried out, is one crucial component of an effective reading program. As part of the Reading First program at the LCO School, two assessments were used for these purposes - the DIBELS ORF and the Reading First SAT. This study is seeking to find out if there was a correlation in the district between student scores on the DIBELS ORF spring benchmark assessment and the total reading score on the Reading First SAT, confirming the use of the ORF as a valid indicator for identifying at-risk readers and documenting their progress.

Subjects

The subjects were 190 students enrolled in the 1^{st} – 3^{rd} grades at the LCO School, located in northwest Wisconsin, from 2004 – 2010. All of the children were Native American and lived within the boundaries of the federally recognized Lac Courte Oreilles Ojibwe reservation.

There were a variety of family/living arrangements that the children lived in - single parent homes headed by mothers; homes headed by a guardian (typically a female relative); homes where there was an unmarried partner living with their parent or guardian; homes where extended family members lived in the same household; and blended families, which included step siblings and half-siblings. In addition, all of the students at the LCO School were related to each other either directly by blood, or indirectly, referring to close acquaintances as “cousins” or “aunties”.

All of the students participated in the free and reduced federal school lunch program. By virtue of their low socio-economic status, all of the students in the study can be considered as

part of the "at-risk" category of early readers – at risk for developing or already having deficits in one or more areas of the reading process - phonemic awareness, phonics, fluency, vocabulary, or comprehension (Willingham, 2012)). Based on their DIBELS assessment recommended proficiency levels, students needing all levels of assistance were included: 1) intensive – needing substantial assistance; 2) strategic – needing extra assistance; and 3) benchmark – needing little/no extra assistance. Subjects included 92 boys and 98 girls, as well 15 readers who were already identified with reading disabilities, having Individual Education Plans developed and in place, and 175 who had not been identified.

Sampling Procedure

All 1st-3rd grade students who attended the LCO School from 2004 – 2010 and who took both the spring DIBELS ORF as well as the Reading First SAT each year were included. They were the ones directly impacted by the reading instruction, intervention, and assessments being carried out as part of the Reading 1st program implemented during those years. Due to the mandates of the Reading First program, all classroom teachers were expected to provide high quality, evidence-based, tier 1 instruction, using the Houghton-Mifflin basal program with fidelity. An uninterrupted, daily ninety minute reading block was established for all K-3rd grade classrooms, with interventions taking place three – five times per week for 30 – 45 minutes per session. Tier 2 and 3 interventions were planned, implemented, and evaluated by a collaborative process between classroom teachers, specialists, para-professionals, reading coach, administration, and parents. DIBELS assessments were used three times yearly for benchmark purposes and every few weeks for progress monitoring. That data, along with any other pertinent testing administered in the classroom was used in a response to intervention, early prevention approach.

Research Setting

The K-12th grade LCO School was located on the Lac Courte Oreilles Ojibwe reservation, in northwest Wisconsin. The school was federally funded and administered through the Bureau of Indian Education (BIE), which is part of the United States Department of Interior. The BIE headquarters were in Washington, DC and Albuquerque, New Mexico, with Office of Indian Education Program (OIEP) technical assistance centers located in each region of the country, the LCO School being served through the Minneapolis/St. Paul office during the years of the study. The state of Wisconsin supplied no direct student aid funding to operate the school. Approximately 300 – 350 students were enrolled at the LCO School, established in 1981 to serve Native American children and provide them with a traditional K-12 "American" curriculum, as well as offering classes and programs in the Ojibwe culture and language.

The DIBELS ORF benchmark tests were administered to the students outside their regular classrooms, in conference rooms located in the office area of the elementary school. Trained, external evaluators (employees from EndVision Research and Evaluation for both the fall and spring assessments; a reading coach from a nearby Native American school for the winter assessments) were contracted with through the Reading First office of the BIE. The Reading First SATs were administered to the students in their regular classrooms, by their classroom teachers. No special accommodations were made in the classrooms, other than spreading the children out around the room as much as possible during testing times.

Research Design

I used a Data Collection "Only" design and looked at two variables – the ORF spring benchmark test scores and the Reading First SAT scores. The tests are the independent variable, with student scores being the dependent variable. I investigated the test scores from six school years: 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2008-2009, and 2009-2010.

Instrumentation

First, I examined the results from the DIBELS ORF benchmark testing (Good and Kaminski, 2003). This assessment was administered to all LCO students in the 1^{st} – 6^{th} grades, three times yearly: 1) fall – in early September; 2) winter – in mid December; and 3) spring – in early May. For the purposes of this study, only the spring results for the 1^{st} – 3^{rd} grade students from six school years, 2004 – 2010, were considered.

The DIBELS ORF spring benchmark assessment consisted of three one-minute timed readings. Both narrative and expository genres were represented in the texts presented to the students. Students were informed that both reading speed, in correct words per minute (cwpm), and accuracy, percentage of words read correctly out of total words attempted, were being tested. The median score was recorded as each student's final score. The test administrator recorded decoding errors using a hand-held computer palm-pilot. This device also kept track of elapsed time, as well as administrator wait time (3 seconds) for a student to decode an unknown word, before prompting the student to go on. Only words deleted or changed by the student were marked wrong – word additions were ignored. If a student skipped a line, it lowered their accuracy, but it wasn't counted against them in calculating their rate/speed. The results were provided instantly for both the administrator and student to view, and they were shown at the discretion of the administrator. After "syncing" the results electronically to the secure website of

LCO School's data management provider, Wireless Generation, the scores were then communicated to staff members and parents in a variety of graphic formats.

The second measure used was the *Reading First Stanford Achievement Test: Tenth Edition* (Harcourt Educational Measurement, 2004). This standardized achievement test was administered to all LCO 1st – 3rd grade students in early May of each school year. The tests were administered by each classroom teacher to her group of students. The subtests included: 1) Word Study Skills – twenty-four questions, covering structural and phonetic analysis; 2) Listening Vocabulary - nine questions, covering receptive oral vocabulary; 3) Reading Vocabulary - nine questions, covering synonyms, multiple meaning words, and context clues; and 4) Reading Comprehension – eighteen to thirty questions (depending on the grade level) covering literal and inferential understandings. The testing format was entirely multiple choice, with three or four response choices for each question. The tests were untimed and administered over a period of two to three days, to minimize testing fatigue.

Treatment of Data

The raw student data from all students included in the study was recorded onto a series of three column charts using the spreadsheet function contained within Microsoft Excel software. Each chart had data from one grade level for one particular year. Student identifying numbers (coded to maintain anonymity) were entered into the first column, SAT scaled scores into the second, and ORF spring benchmark scores into the third. The Pearson Product Moment Correlation formula, contained within the Excel software program, was then used to compute the correlations between the SAT scaled scores and ORF spring benchmark scores for each grade level and year. Scatterplot graphs were created next, again using the Excel software. Ovals were

drawn around the area on each scatterplot graph where most of the data points were concentrated. Finally, that correlation data was recorded onto a four column chart, using the spreadsheet function contained within Microsoft Excel software. The first column listed the school year, with the last three columns containing the correlation data from each grade level, first - third.

Validity and Reliability of Instrumentation

The ORF and SAT tests were selected because the LCO School was mandated to use both assessments, as part of the BIE Reading 1st guidelines agreed upon when funding was received. Both tests have been administered to thousands of students in almost every state for a number of years. Researchers for both tests have developed standardized norms for the interpretation of student results in comparison to grade level peers. Both testing companies also engage in an ongoing process of updating norms, taking into account changing student characteristics and conditions under which the tests are administered. Both are also criterion-referenced assessments, providing information about a student's level of achievement in relation to established standards.

The DIBELS ORF benchmark tests were developed by The Early Childhood Research Institute on Measuring Growth and Development (ECRI-MGD) at the University of Oregon. Their literature states that research has, "validated its ability to predict outcomes, and tested its reliability using data from thousands of young children in many regions of the country" (Good & Kaminski, 2002). The ORF test, developed originally in the late 1970's and field tested beginning in the 1980's, was, at first, not uniformly administered as a one minute timed test. These early researchers eventually concluded that administering the ORF for one minute was,

"just as reliable, valid, and indicative of student skill as information obtained in multi-minute samples" (Coulter et al. 2009). They also found that when three passages were administered to a student, instead of just one, ORF reliability is increased (Good & Kaminski, 2002).

The Reading First SAT test was developed in response to the mandates of NCLB, as well as the report of the National Reading Panel (NRP), which called for the use of scientifically-based reading research (SBRR) best practices for the teaching of reading. It took selected test items and sections from the existing Stanford Achievement Test Series to gauge students' proficiency in the foundational early reading skills. Questions were selected to measure student progress in the five essential components of reading identified by the NRP: phonemic awareness, phonics, fluency, vocabulary, and comprehension.

One way the results were reported was in a norm-referenced format, using indicators such as scaled score, national percentile ranking, and national curve equivalent. The results were also reported in a criterion-referenced format, with levels of proficiency indicated for each tested area as well as for the entire test. Three levels of reading proficiency were delineated: 1) Needs Substantial Intervention; 2) Needs Additional Intervention, and 3) At Grade Level. The test publisher also stated in its literature, "*Stanford Reading First* is a technically excellent, valid, and reliable instrument with 2002 norm-referenced information (Harcourt Educational Measurement, 2003)."

Methods of Data Analysis

The *Pearson Product-Moment Correlation* formula was used to determine the relationship between the ORF spring benchmark scores and Reading First SAT scores. The computations involved in the formula were performed using Microsoft Excel spreadsheet

software. The data was analyzed by grade level for each individual year, and overall, for each grade level, over the four school years (2004 – 2010). The data was examined in aggregate, covering all grades, 1st – 3rd and all years, 2004 – 2010. All results were displayed on scatterplot graph(s), which were evaluated, looking at the distribution of the points as well as significance of any outliers.

Rationale for Methods of Data Analysis

The *Pearson Product-Moment* formula was used to analyze the data because the relationship between a student's score on the DIBELS ORF test and Reading First SAT Total Reading Score was being examined.

Summary

For struggling readers to achieve success, practitioners need to first be able to identify and prioritize those students most in need of interventions. An effective method of evaluating how students are progressing in relation to the evidence-based instruction being carried out is also needed. This study examines two assessments the staff at the LCO School used for these purposes. It seeks to find out if the DIBELS ORF screening assessment, a test to help identify at-risk readers based on one part of the reading process, fluency, is a good predictor of overall reading achievement, as measured by the Reading First SAT.

Chapter 4

Introduction

One important component of a district's effective reading program is the use of valid and reliable assessments that help inform instructional decisions, identify children in need of intervention, and gauge the results of the research-based instruction being carried out (National Research Council, 1998). Two assessments used as part of the Reading 1st program at the LCO School were the DIBELS ORF test and the Reading First SAT. This study sought to find out if there was a correlation between a student's score on the DIBELS ORF spring benchmark assessment and Total Reading Score on the Reading First SAT.

Data Presentation

The correlations relating to the LCO 1st – 3rd grade student results on their DIBELS spring ORF and Reading First SAT assessments from 2004 – 2010 are shown in Figure 1 of the appendix.

The 1st grade results for each of the six years, as well as for the years in aggregate, all showed significant positive correlations. The range of all scores was r = .30, from a low of r = .52 for the 2009-2010 school year, to a high of .82 for the 2005-2006 school year. The mean for all six years was r = .71 and the median was r = .67.

The 2nd grade results for each of the six years, as well as for the years in aggregate, showed significant positive correlations, except for 2004-2005, which displayed a value of r = .38. The range of all scores was r = .31, from a low of .38 for the 2004-2005 school year, to a high of .69 for the 2008-2009 school year. The mean for all six years was r = .56 and the median was r = .54.

The 3rd grade results for each of the six years, as well as for the years in aggregate, showed mixed results. Three of the years, 2004-2005, 2005-2006, and 2006-2007, as well as the years in aggregate, showed significant positive correlations, r = .53, r = .66, r = .54, and r = .45. Two of the years, 2008-2009 and 2009-2010, displayed positive correlations, although at lower levels, r = .24 and r = .37. The result for the 2007-2008 school year showed almost no correlation, r = .04. The range of all scores was r = .62, from a low of r = .04 for the 2007-2008 school year, to a high of r = .66 for the 2005-2006 school year. The mean for all six years was r = .40 and the median was r = .35.

The correlations relating to the LCO 1st – 3rd grade student results on their DIBELS spring ORF and Reading First SAT assessments from 2004 – 2010 are displayed on scatterplot graphs in Figures 2 – 22 of the appendix. Each graph shows the configuration, slope, and clustering of correlation values represented by the data. Ovals encircle a majority of the data points on each graph, to help evaluate the effects outlying data points and to more easily view the slope and configuration.

The scatterplot graphs representing the 1st grade correlations from 2004 – 2010 are shown in Figures 2 – 8 of the appendix. In six out of the seven graphs, covering the years 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2009-2010, and 2004-2010, there are no significant outliers. So the results from these years can be considered to be true representations, not unduly skewed by data that is a substantial distance away from the norm. Only one year, 2008-2009, had a significant outlier which could exert an undue effect on the correlation, although if the slope of the existing oval is extended, which may more accurately reflect its true trend, that data point becomes only marginally significant.

The scatterplot graphs depicting the 2nd grade correlations from 2004 – 2010 are shown in Figures 9 – 15 of the appendix. None of the graphs have any data points that are significantly outside the norm, so those outcomes can be regarded as accurate.

The scatterplot graphs representing the 3rd grade correlations from 2004 – 2010 are shown in Figures 16 – 22 of the appendix. Four out of the seven graphs contain data points which are a substantial distance away from the norm. 2006-2007 and 2008-2009 each have one significant outlier, 2009-2010 has two, and 2007-2008 has one significant outlier and one moderately significant one. 2004-2005 had one data point that was a moderate distance away from the norm and two years, 2005-2006 and 2004-2010, had no data points substantially away from the norm that could skew the results.

Summary

The purpose of this study was to determine if a student's score on the DIBELS ORF spring assessment correlates with reading progress as shown on the Reading First SAT. The results for each grade level for the individual years 2004-2010 demonstrated positive correlations all six 1st grade years, all six 2nd grade years, and five out of six 3rd grade years, for a total of seventeen out of the eighteen years examined. Additionally, fourteen out of the eighteen years demonstrated significantly high correlations. Regarding the aggregate years 2004 - 2010 investigated for each grade level, all three grades displayed significantly high correlations.

Chapter 5

Introduction

This study supports the use of DIBELS ORF for early identification of at-risk readers at the LCO School in concert with the instructional guidelines in place as a result of participation in the Reading First program. The ORF can be part of a comprehensive student measurement system where decisions are driven by the use of valid assessments in an RTI framework. The information gained from the ORF can be used to make adjustments to instruction and intervention for both individuals and groups of students.

Deductive Conclusions

For LCO students in 1st – 3rd grades, from 2004 – 2011, the data shows a positive correlation between a student's score on the DIBELS ORF spring benchmark assessment and his/her Total Reading Score on the Reading First SAT. The strength of the correlations is medium to fairly strong as indicated by the narrow shape of the ovals that encircle the values on the scatterplot graphs. There were few significant outlying scores that complicated the interpretation of the data. Also, there was no atypical restriction or extension of the range of values that would affect the interpretation of the data. A teacher can be confident that a student's DIBELS ORF screening test score is an valid indicator of his/her overall reading ability and take steps to provide that student with the most appropriate and effective instruction, in part based on that score.

Discussion of Data

The strong, positive correlation between the DIBELS ORF scores and the Reading First SAT scores for LCO students confirmed what previous researchers have found - the DIBELS ORF test is a valid indicator for early identification of at-risk readers and documentation of their progress.

Both Cook (2003) and Baker et al. (2008) found significant correlations between DIBELS and SAT scores. For Cook's subjects, a group of rural Ohio 1st graders, the correlation was (r = .73). Baker reported a correlation of (r = .82) for a 1st grade group from Oregon and (r = .80) for a 2nd grade group. Various researchers have also found high correlations between the DIBELS tests and state assessments for 3rd grade groups. The following results were reported examining the relationship between DIBELS and each assessment listed: 1) North Carolina Reading Assessment (r = .73) (Barger 2003); 2) Oregon Statewide Reading Assessment (r = .67) (Good, Simmons, & Kame'enui, 2001); 3) Colorado State Assessment Program (r = .80) (Shaw & Shaw, 2002); 4) Arizona Instrument to Measure Standards (r = .74) (Wilson, 2005); and 5) Florida Comprehensive Assessment Test (r = .70) (Buck & Torgesen, 2002). Pressley and his associates (2005) found a positive correlation between DIBELS and results on the Terra Nova assessment, although to a lesser degree, with a 3rd grade group of Midwestern urban students (r = .45). Finally, for a 1st grade group of students, Elliott et al. (2001) found the correlation between DIBELS and scores from the Woodcock-Johnson Broad Reading Cluster to be high (r = .76).

In researching the validity of ORF for identifying at-risk readers and documenting their progress, Baker and his collegues (Baker et al., 2008) stated that all 34 Oregon Reading 1st schools in their study were providing for their students highly specific and prescribed methods of

reading instruction – the use of DIBELS screening assessments being among them. As a result, they conclude that a great deal is known about the instructional setting which brought about the growth in student achievement. The use of DIBELS assessments was one important component of the total reading programs in these schools that contributed to higher levels of student achievement.

The LCO students in this study, who were also part of the Reading First initiative, were taught under similar instruction conditions. For K-3rd graders, there was a daily ninety minute minimum uninterrupted reading block. During that time instruction was focused on the essential elements of beginning reading (National Reading Panel, 2000) using scientifically based reading researched methods. Student needs, as determined by DIBELS screening and progress monitoring data, drove decisions concerning the emphases of instruction such as group size, membership, and materials selection. Intervention instruction outside the ninety minute reading block was provided for identified at-risk readers targeting their deficient skill areas. The progress of all students was measured through the administration of three DIBELS benchmark assessments, regular DIBELS progress monitoring, and the Reading First SAT each year. Staff teams used this data to analyze, plan, and adjust instruction. A Reading Coach supported teachers and teams in the implementation of effective instruction and as part of ongoing, high quality professional development.

Implications

The DIBELS assessments in Reading First schools were used not only three times yearly for the screening and identification of at-risk readers, but also 1 – 3 times monthly for the purposes of progress monitoring and adjusting instruction to meet student needs. Baker and his

colleagues concluded that data gained from DIBELS progress monitoring assessments would likely correlate with the benchmark data (Baker et al., 2008), providing further support and validation for the use of those tests as well at the LCO School.

In a district seeking to operate under the principles of early identification of struggling readers and Response to Intervention, school psychologists play an important role. They need to help staff use screening and progress monitoring data to make decisions about the effectiveness of interventions for individuals and groups of students (Shinn et al., 2002). In order for all students to make progress in meeting key learning goals, systems need to be developed integrating effective, appropriate instruction and intervention with the use of accurate measurements of learning. School psychologists need to be leading the way to assist staff in this complex process.

Because research-based instruction and intervention needs to be implemented in a systemic fashion for it to be successful, looking at data not only on the individual level, but also on group levels, looking for patterns and trends, will be important. School psychologists, again, are the ones most able to meet this need. For example, if the data is showing a small number of students consistently performing poorly, even though the instruction and interventions have been found to be effective, the focus for future decisions would seem to call for adjusting the instructional program and interventions for those specific students. If, on the other hand, numerous students are struggling in spite of verified, effective instruction and intervention, it could well indicate the source of the problem is systemic and adjustments need to be made on this "macro" level. The success of a district's implementation of RTI depends on this type of problem solving (Batsche et al., 2005; Tilly, 2008; IDEA, 2004).

These principles of early identification/prevention and RTI, using the DIBELS ORF screening test as a valid indicator for early identification of at-risk readers and documentation of their progress, also have implications for special education staff. In the process used to identify a learning disability and qualify a student for special education services, RTI has been identified as an alternative (Fuchs and Fuchs, 1998; Vaughn et al., 2003; IDEA, 2004). How schools measure student progress in response to evidence-based interventions, in determining whether it is sufficient or not, is a crucial part of more closely integrating general and special education instruction (Gersten & Dimino, 2006; Fuchs & Fuchs, 2006). Students who continue to demonstrate less than adequate progress to scientifically based reading researched instruction and intervention are then evaluated for the possible presence of a learning disability. In this process, because of the importance placed upon valid, reliable screening and progress monitoring measures to identify at-risk students and evaluate their response to intervention in determining eligibility for placement in special education, school psychologists again play a key role. They need to help their colleagues implement sound, evidence-based instructional and assessment practices (Messick, 1989; Gersten & Dimino, 2006; Fuchs and Fuchs, 2006).

For Further Research

The sample of students for this study was limited to one district, the Lac Courte Oreilles Ojibwe School. Replication of the study is needed with a broader sample of students to establish how well the findings can generalize to other schools and districts. Districts that participated in the Reading 1st initiative, such as the LCO School, were chosen due to having high percentages of their student populations coming from low socio-economic families. Schools eligible to apply for the grant also needed to demonstrate that previous years' instruction resulted in low reading

achievement. The findings of this study are very likely able to be generalized to other schools having these characteristics. In addition, these findings are likely comparable to schools who didn't take part in the Reading First program, although, at this time, that is not known. Future research should also investigate the relationship between DIBELS ORF scores and high stakes testing regarding students in grades other than $1^{st} – 3^{rd}$, as well as with distinct student populations: students placed in special education programming, students from middle and high socio-economic families, students attending non-public schools (private, voucher, on-line), ELL and non-ELL students, and specific demographic groups (Hispanic, Asian, Negro, etc.).

Summary

The use of DIBELS ORF for screening and identification of struggling readers, within the Reading First initiative's context of scientifically-based reading researched methods of instruction and intervention for the prevention and remediation of reading difficulties, has been supported by this study. As part of a comprehensive assessment system that districts develop to make decisions on individuals and groups of students, the DIBELS ORF test can be an integral part of the process. It can provide timely and useful information to school staff about which students are on a trajectory towards acceptable reading achievement and which ones aren't, leading to effective adjustments to instructional programming. The growth students demonstrate on ORF over time can be a valuable gauge in showing not only how reading fluency is developing, but also other skills that contribute to the overall process of reading.

Appendix

Figure 1

2004-2010 1st - 3rd Grade ORF-SAT Correlations

	1st Grade	2nd Grade	3rd Grade
2004-2005	0.67	0.38	0.53
2005-2006	0.82	0.67	0.66
2006-2007	0.76	0.44	0.54
2007-2008	0.77	0.49	0.04
2008-2009	0.69	0.69	0.24
2009-2010	0.52	0.68	0.37
2004-2010	0.69	0.57	0.45

Figure 2 (r = .67)

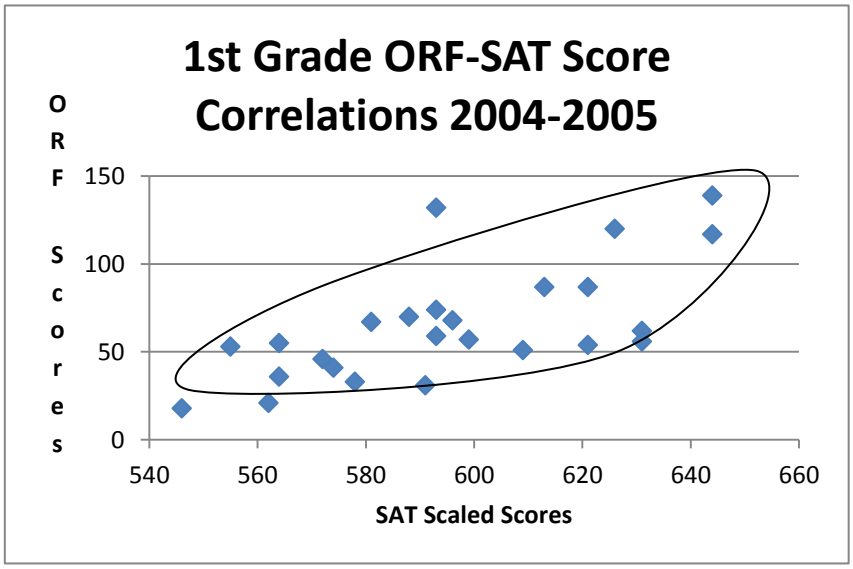


Figure 3 (r = .82)

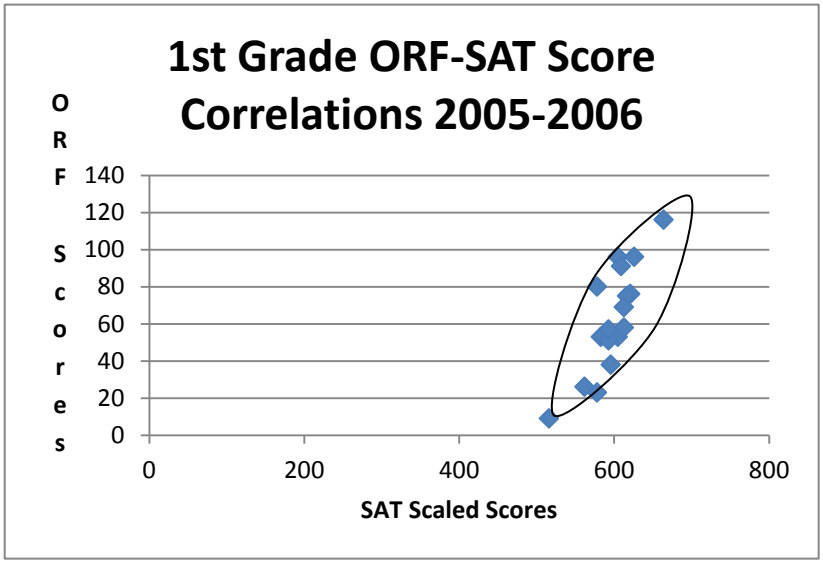


Figure 4 (r = .76)

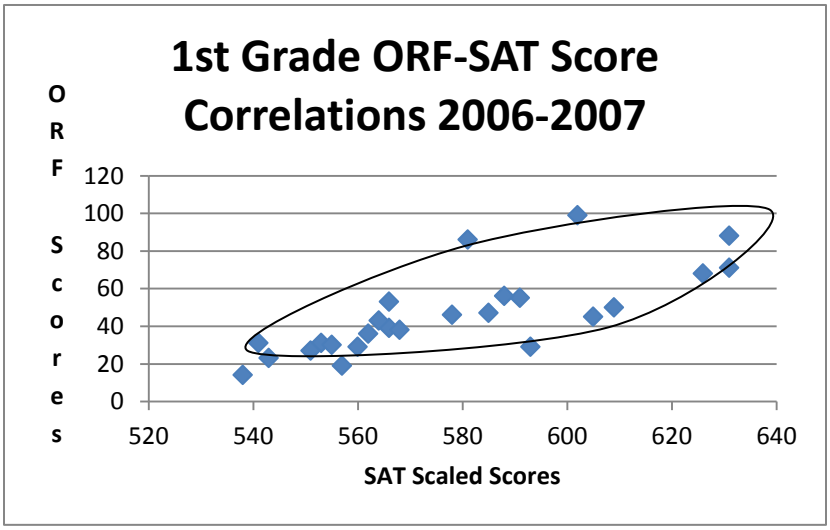


Figure 5 (r = .77)

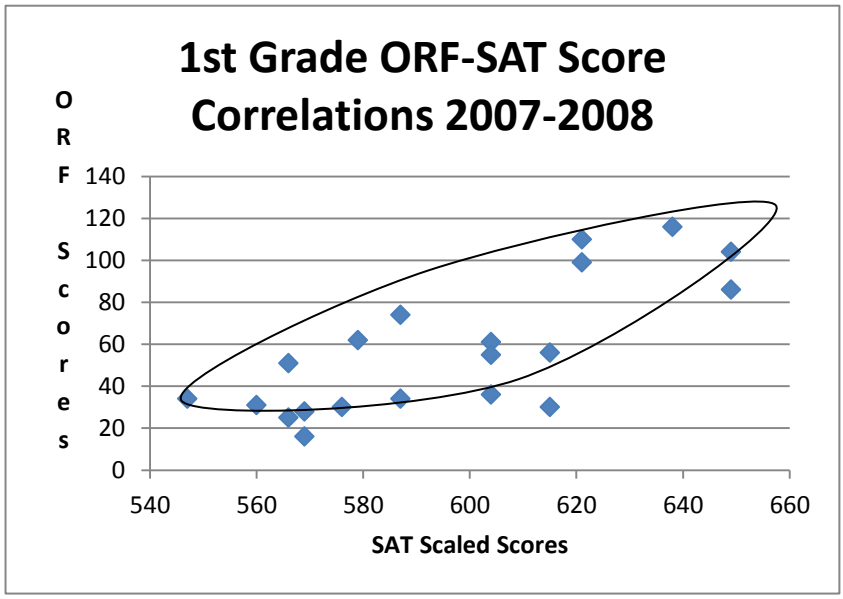


Figure 6 (r = .69)

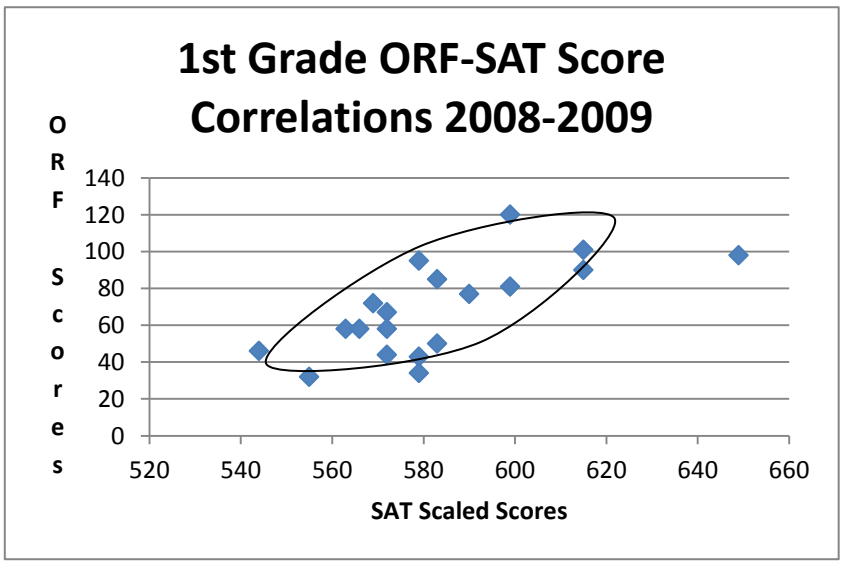


Figure 7 (r = .52)

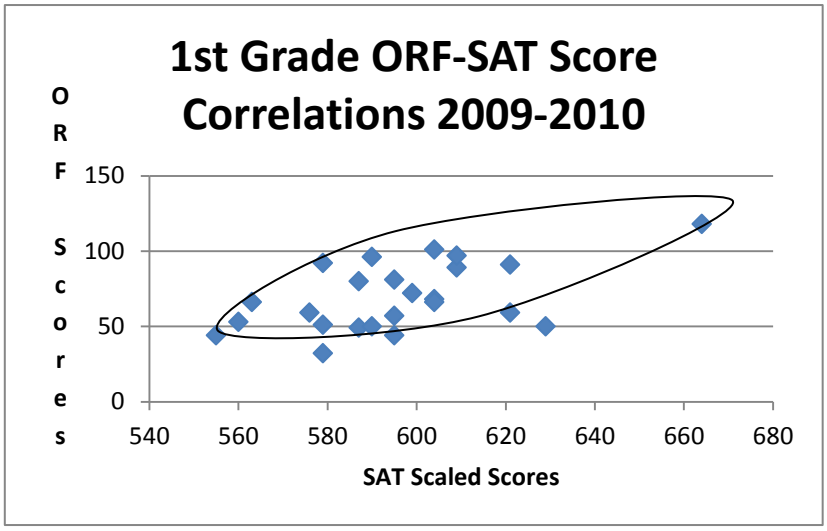


Figure 8 (r = .69)

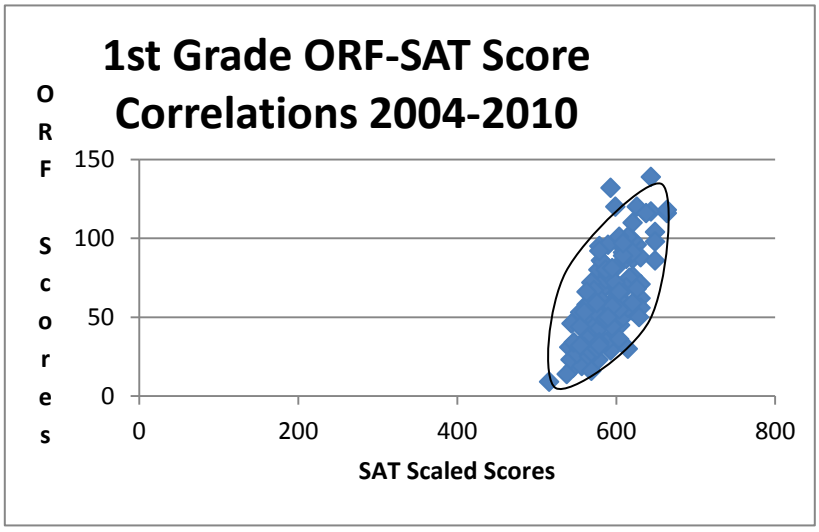


Figure 9 (r = .38)

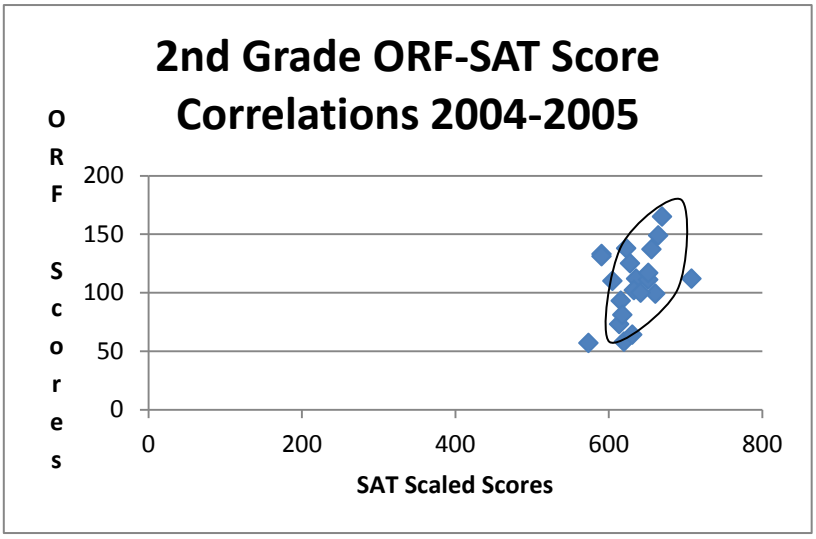


Figure 10 (r = .67)

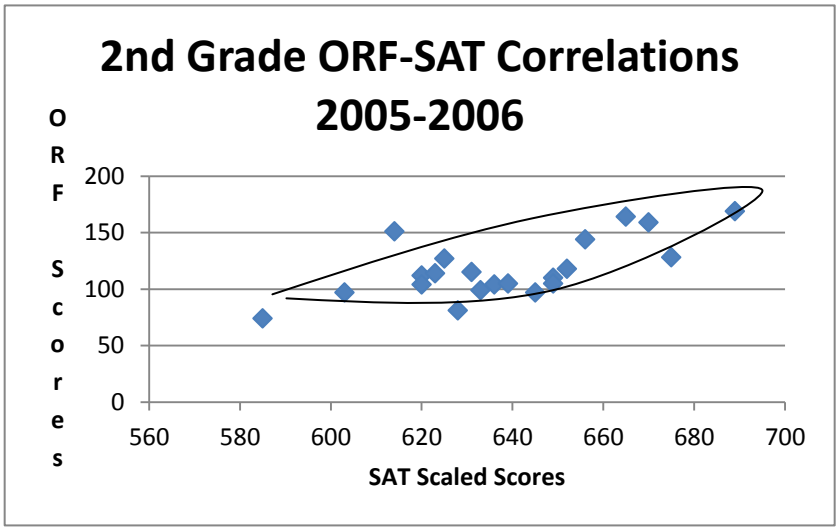


Figure 11 (r = .44)

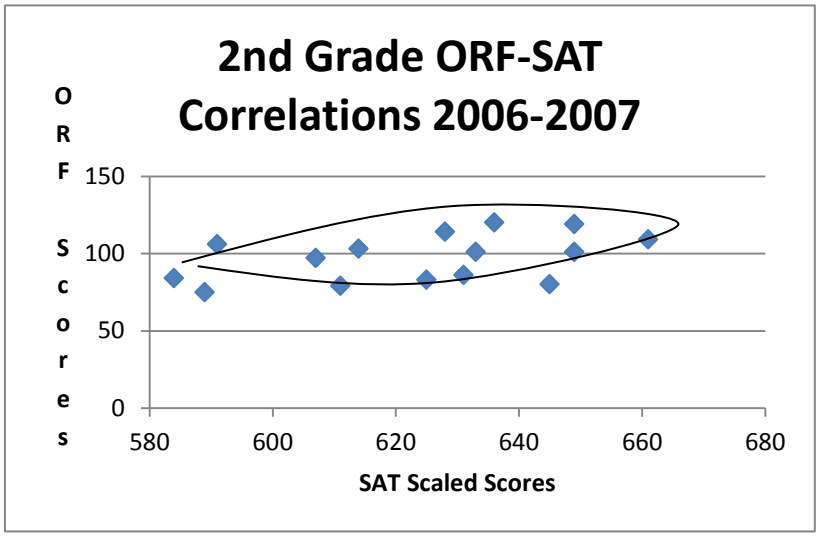


Figure 12 (r = .49)

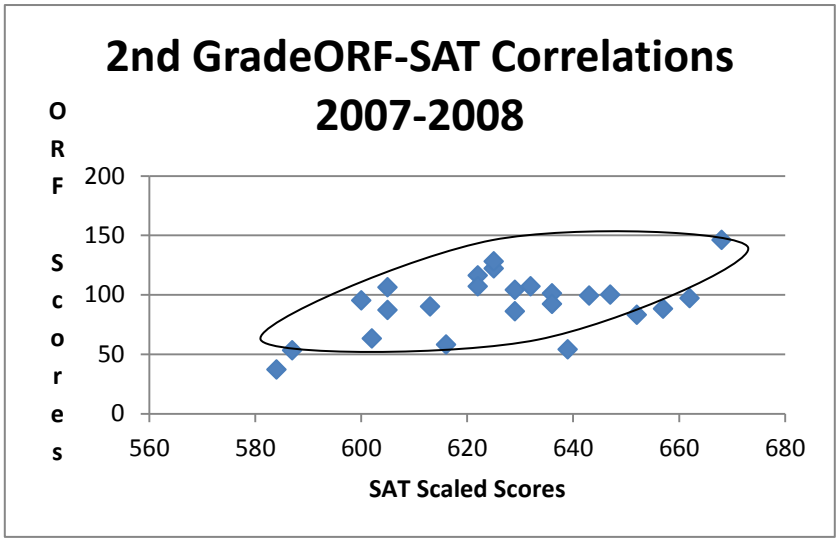


Figure 13 (r = .69)

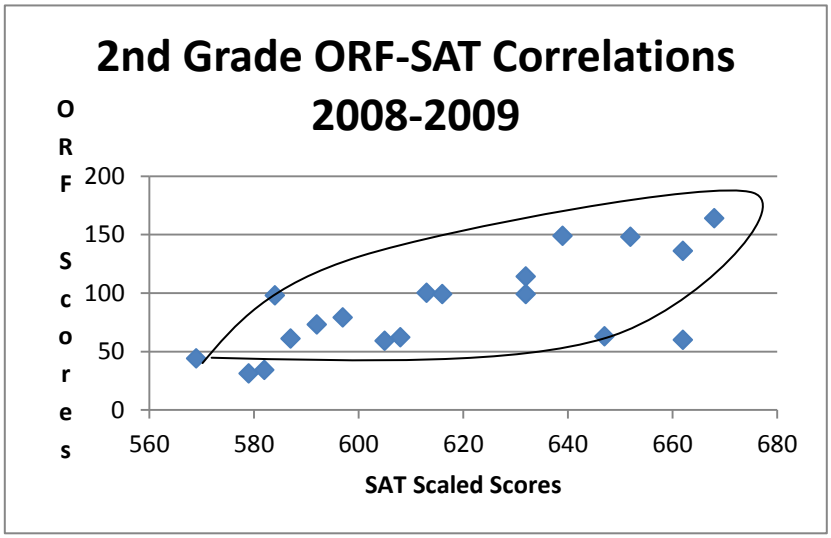


Figure 14 (r = .68)

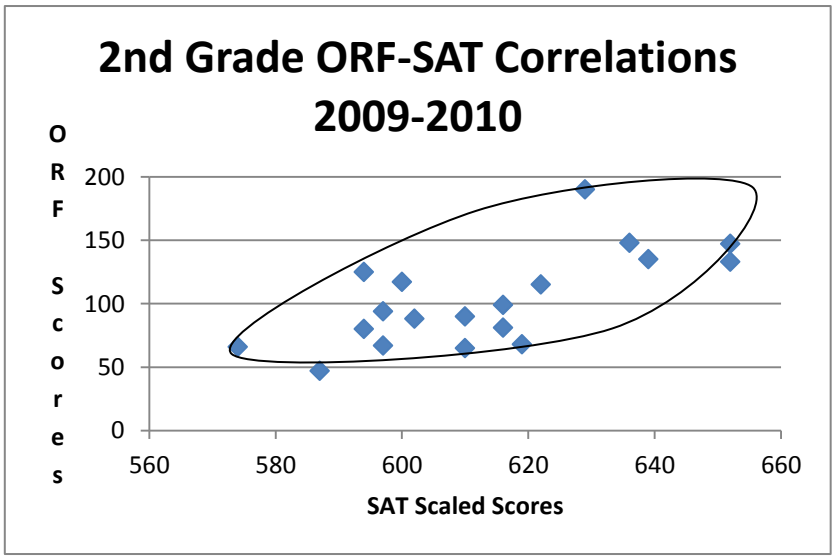


Figure 15 (r = .57)

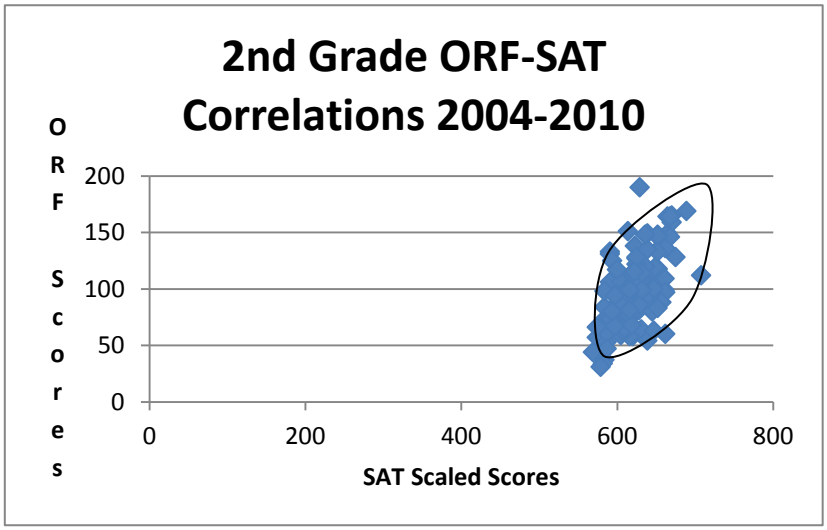


Figure 16 (r = .53)

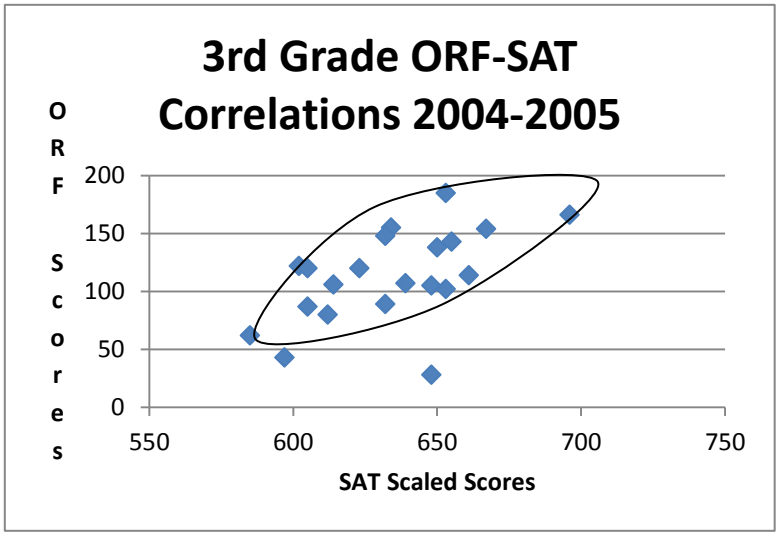


Figure 17 (r = .66)

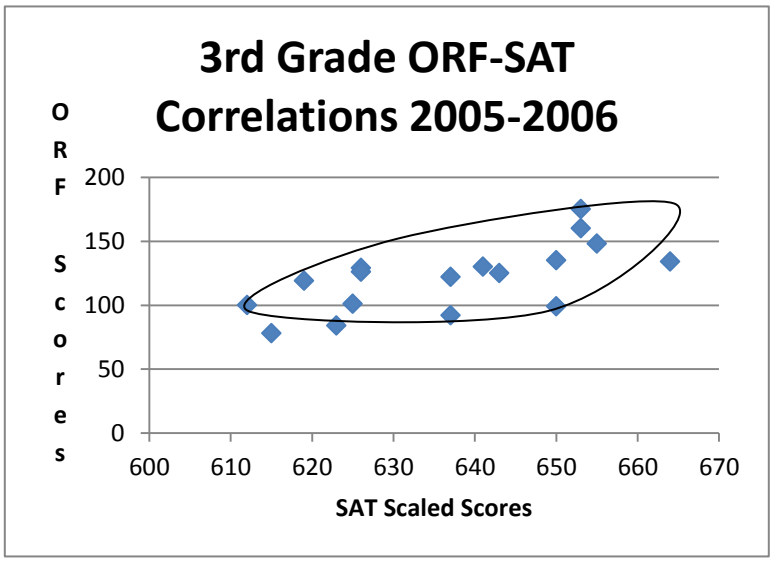


Figure 18 (r = .54)

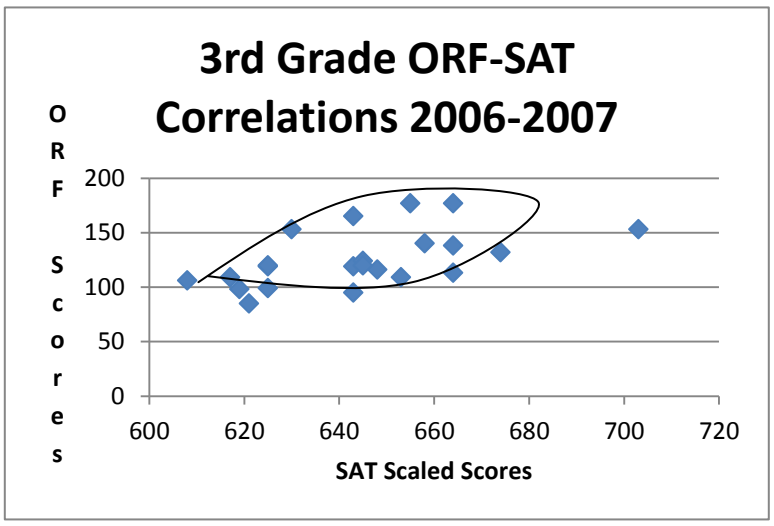


Figure 19 (r = .04)

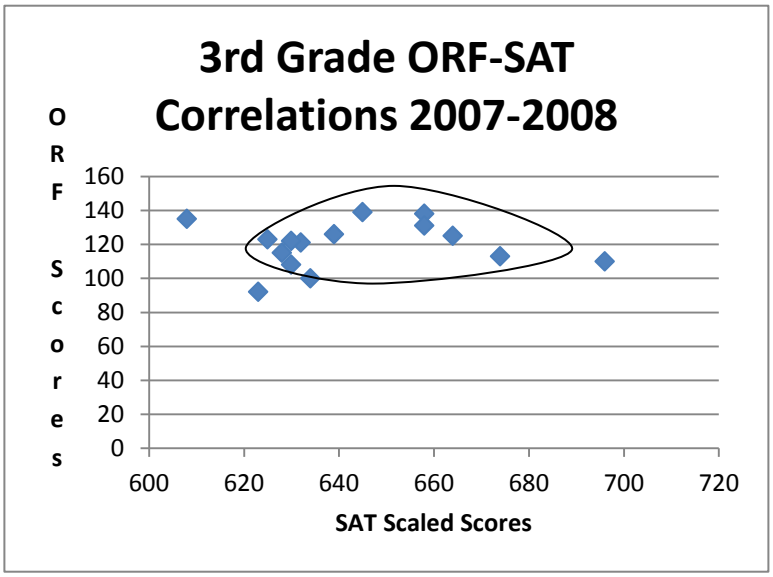


Figure 20 (r = .24)

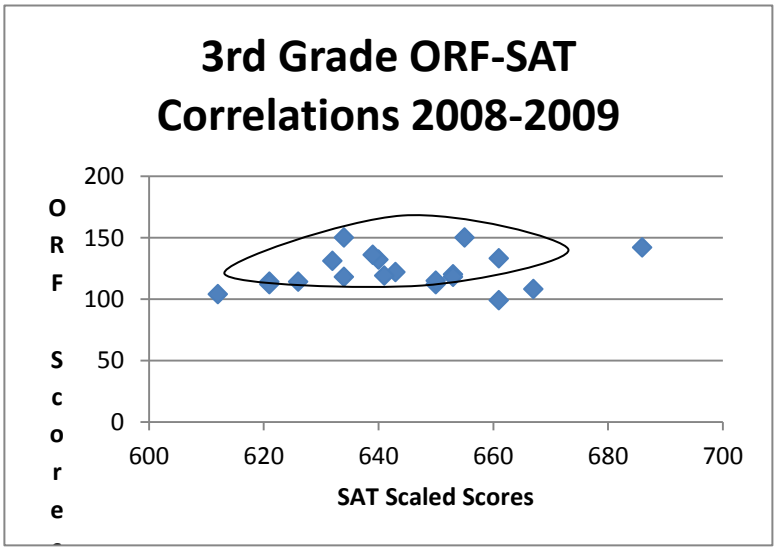


Figure 21 (r = .37)

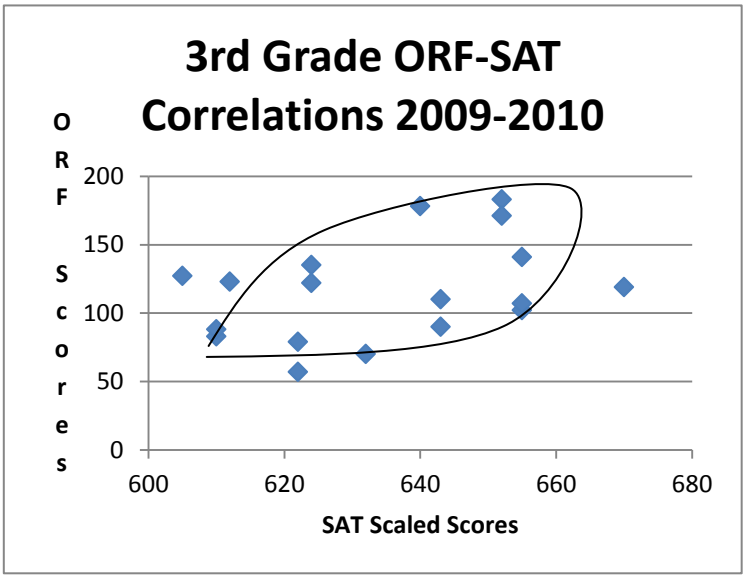
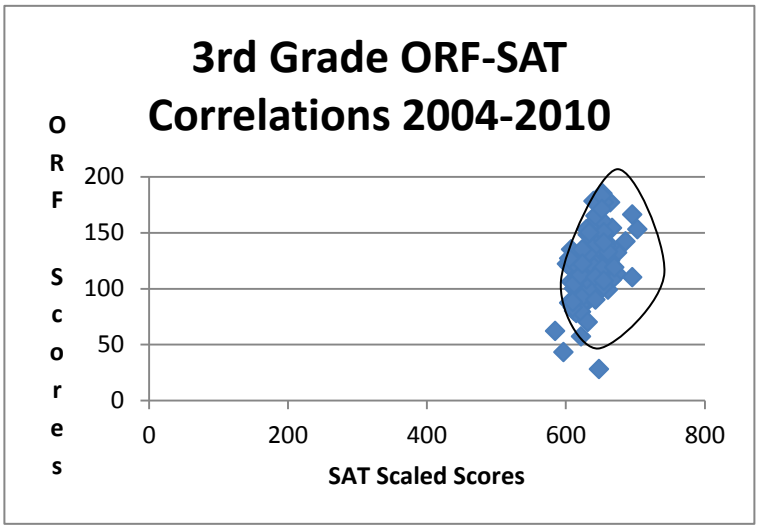


Figure 22 (r = .45)



References

Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Boston, MA: Colliers.

Al Otaiba, S., & Fuchs, D., (2002). Characteristics of children who are unresponsive to early literacy intervention. *Remedial and Special Education. 23(5),* 300-316.

Atkinson, B. (2012). Target practice: Reader response theory and teachers' interpretations of students SAT 10 scores in data-based professional development. *Journal of Teacher Education*. 63 (3), 201-213.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E., & Beck, C. T., (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*. 37 (1), 18-37.

Barger, J. (2003). *Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Asheville, NC: North Carolina Teacher Academy.

Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovaleski, J., Prasse, D., et al. (2005). IDEA '04. *Response to intervention: Policy considerations and implementation.* Alexandria, VA: National Association of State Directors of Special Education (U.S.).

Brenner, D. & Hiebert, E. (2010). If I follow the teachers' editions, isn't that enough? Analyzing reading volume in six core reading programs. *Elementary School Journal*, 110 (3), 347 – 363.

Buck, J. & Torgesen, J. (2002). The relationship between performance on a measure of oral reading fluency and performance on the Florida comprehensive assessment test. (Technical Report) Tallahassee, FL: Florida Center for Reading Research.

Center for the Improvement of Early Reading Achievement. (2001). *Put reading first: The research building blocks for teaching children to read.*

Chard, D. J., & Kame'enui, E. J. (2000). Struggling first-grade readers: The frequency and progress of their reading. *Journal of Special Education*, 34, 28-38.

Chard, D. J., Stoolmiller, M., Harn, B. A., Wanzek, J., Vaughn, S., & Linan-Thompson, S. (2008). Predicting reading success in a multilevel schoolwide reading model. *41(2),* 174-188.

Cook, R. G. (2003). The utility of DIBELS as a curriculum based measurement in relation to reading proficiency on high stakes tests. (Thesis). Marshall University Graduate College.

Coulter, G., Shavin, K., & Gichuru, M. (2009). Oral reading fluency: Accuracy of assessing errors and classification of readers using a 1-minute timed reading sample. *Preventing School Failure*, 54 (1), 71 – 76.

Dessoff, A. (August 2007). DIBELS draws doers & doubters. *District Administration*, August 2007, 38-43.

Dwyer, K., Osher, D., & Hoffman, C.C. (2000). Creating responsive schools: Contextualizing early warning, timely response. *Exceptional Children*, 66, 347-365.

Elliott, J., Lee, S., & Tollefson, N. (2001). A reliability and validity study of the dynamic indicators of basic early literacy skills – modified. *School Psychology Review*, 30 (1), 33.

Euneau, C. (2004 – 2010). Annual yearly progress records at Lac Courte Oreilles Ojibwe School.

Fuchs, L. S., & Fuchs D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice*, 13, 204-219. Current issues in special education and reading instruction-Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41 (1), 92.

Fuchs, D., & Fuchs, L. S. (2006). Current issues in special education and reading instruction-Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41 (1), 92.

Gersten, R., & Dimino, J. A. (2006). RTI (response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly*, 41 (1), 92.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: .

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*. 5 (3), 257-288.

Good, R .H., Simmons, D., Kame'enui, E., Kaminski, R., & Wallin, R. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade.* Eugene, OR: Center for Improving Reading Competance Using Intensive Treatements Schoolwide: Technical Report #11.

Goodman, K. (2007). One minute of nonsense: The truth about DIBELS. *District Administration*, August 2007, 42-43.

Haager, D., Klingner, J., & Vaughn, S. (2007). *Evidenced-based reading practices for response to intervention*. Baltimore: Brookes.

Hale, J. (2001). *Learning while Black: Creating educational excellence for African American children*. Baltimore, MD: Johns Hopkins University Press.

Harcourt Assessment. (2004). *Stanford Achievement Test Series, Tenth Edition*. Pearson Education, Inc. San Antonio, TX.

Hecht, S. A. & Greenfield, D. B. (2001). Comparing the predictive validity of first grade teacher ratings and reading-related tests on third grade levels of reading skills in young children exposed to poverty. *School Psychology Review*, 30 (1), 50-69.

Individuals With Disabilities Education Improvement Act of 2004, Pub. L. 108-466-614, Stat. 2706 (2004).

International Reading Association. (2008). *IRA Commission on RTI: Working draft of guiding principles*. RTICommission@reading.org.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, 35 (7), 3-12.

Landsman, J. (2004). Confronting the racism of low expectations. *Educational Leadership*, 62 (4), 28-32.

Lynn, M., Bacon, J., Totten, T., Bridges, T., & Jennings, M. (2010). Examining teachers' beliefs about African American male students in a low-performing high school in an African American school district. *Teachers College Record*, 112 (1), 289-330.

McKenna, M. C., & Stahl, S. A. (2003). *Assessment for reading instruction*. New York: Guilford Press.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71(4),* 445-463.

Mueller, A. (2004 – 2010). Special education referral records at Lac Courte Oreilles Ojibwe School.

National Institute for Literacy, *Put Reading First: The Research Building Blocks for Teaching Children to Read - Kindergarten Through Grade 3*, 2000, www.nifl.gov.

National Institute of Child Health and Human Development, *Report of the National Reading Panel. Teaching Children to Read: An Evidence-based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction: Reports of the Subgroups*, NIH Publication No. 00-4754 (Washington, DC: U.S. Government Printing Office), 2000.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Pub. No. 00-4769). Washington, DC: National Institute of Child Health and Human Development.

National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Noguera, P. (2007). Extended view: Race, student achievement and the power and limitations of teaching. *Sage Race Relations Abstracts*, 32, 44-47.

Olson, L. (2007).Instant read on reading, in palms of their hands. *Education Week*. *26(35)*, 1-5.

107th Congress (2002). *The No Child Left Behind Act of 2001*. Washington, DC: United States Congress.

Pearson, P. D. (August 2007). DIBELS draws doers & doubters. *District Administration*, 38-43.

Pressley, M., Hilden, K., & Shankland, R. (2005). An evaluation of end-grade-3 dynamic indicators of basic early literacy skills (DIBELS): Speed reading without comprehension, predicting little. Michigan State University.

Shaw, R. & Shaw, D. (2002) DIBELS Oral reading fluency-based indicators of third grade reading skills for Colorado State assessment program (CSAP). (Technical Report) Eugene, OR: University of Oregon.

Shinn, M. R., Shinn, M. M., Hamilton, C., & Clarke, B. (2002). Using curriculum-based measurement in general education classrooms to promote reading success. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Prevention and remedial approaches* (pp. 113-142). Bethesda, MD: National Association of School Psychologists.

Simmons, D. C., Kame'enui, E. J., Stoolmiller, M., Coyne, M.D., & Harn, B. (2003). Accelerating growth and maintaining proficiency: A two-year intervention study of kindergarten and first-grade children at-risk for reading difficulties. In B. R. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 197-228). Timonium, MD: York Press.

Thomas B. Fordham Institute. (2007). *Whole-language high jinks* [Brochure]. Dayton, OH: Louisa Moats.

Thomas B. Fordham Institute. (2000). *Whole language lives on: The illusion of "balanced" reading instruction*. Dayton, OH: Louisa Moats.

Tilly, D. (2008). The evolution of school psychology to science based practice. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 17-36). Bethesda, MD: National Association of School Psychologists.

Torgesen, J. K. (2000).Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*. *15(1)*, 55-64.

Vadasy, P. F., Sanders, E. A., Peyton, J. A., & Jenkins, J. R. (2002). Timing and intensity of tutoring: A closer look at the conditions for effective early literacy tutoring. *Learning Disabilities Research & Practice*, *17(4)*, 227-241.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children*, 69, 391-409.

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Pedrotty B. D., Dickson, S., & Blozis, S. (2003). Reading instruction grouping for students with reading difficulties. *24(5)*, 301-315.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R. S., et al. (1996). Cognitive profiles of difficult to remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, 88, 601-638.

Vellutino, F. R., Scanlon, D. M., Small, S. G., & Fanuele, D. (2003, December). *Response to intervention as a vehicle for distinguishing between reading disabled and non-reading disabled children: Evidence for the role of kindergarten and first grade intervention.* Paper presented at the National Research Center on Learning Disabilities Responsiveness to Intervention Symposium, Kansas City, MO.

Willingham, D. T. (2012). Why does family wealth affect learning? *American Educator*, 36 (1), 33 – 39.

Wilson, J. (2005). *The relationship of dynamic indicators of basic early literacy skills (DIBELS) Oral Reading Fluency to performance on Arizona instrument to measure standards (AIMS).* Research Brief: Assessment and Evaluation Department, Tempe School District No. 3. Tempe, AZ.